

CLOCK SYNCHRONIZED NONVOLATILE MEMORY DEVICE

The present application is a continuation of application Serial No. 10/223,485, filed August 20, 2002; which is a continuation of application Serial
5 No. 10/020,873, filed December 19, 2001; which is a continuation of application Serial No. 09/817,021, filed March 27, 2001, now U.S. Patent No. 6,366,495; which is a continuation of application Serial No. 09/583,949, filed May 31, 2000, now U.S. Patent No. 6,256,230; which is a continuation of application Serial No. 09/287,187, filed April 6, 1999, now U.S. Patent No.
10 6,111,790; which is a continuation of application Serial No. 09/053,494, filed April 2, 1998, now U.S. Patent No. 6,038,165; which is a continuation of application Serial No. 08/860,793, filed July 9, 1997, now U.S. Patent No. 5,889,698, the contents of which are incorporated herein by reference.

15 TECHNICAL FIELD

The present invention relates to a technique which is especially effective when applied to a multi-value data storing system in a semiconductor memory device and a nonvolatile semiconductor memory device, for example, to a technique which is effective when applied to a nonvolatile memory device
20 (hereinafter referred to as the 'flash memory') for batch-erasing a plurality of memory data electrically.

BACKGROUND OF THE INVENTION

A flash memory uses nonvolatile memory elements each having a
25 control gate and a floating gate similar to FAMOSs, as its memory cells, and

each memory cell can be constructed of one transistor. In such a flash memory, for a programming operation, the drain voltage of the nonvolatile memory element is set to about 5 V, as shown in Fig. 12, and the word line connected to the control gate is set to about -10 V, so that the charge on the floating gate is drawn therefrom by tunnel current to set the threshold voltage to a low value (logic "0").

For the erasing operation, as shown in Fig. 13, the P-type semiconductor region pwell is set to about -5 V, and the word line is set to about 10 V, so that tunnel current is caused to flow to inject negative charges into the floating gate, thereby to set the threshold value to a high state (logic "1"). Thus, one memory cell is able to store the data of one bit.

Incidentally, the concept of a so-called "multi-value" memory has been proposed in which data of two or more bits are stored in one memory cell so as to increase the storage capacity. An invention relating to such a multi-value memory is disclosed in Japanese Patent Laid-Open No. 121696/1984.

In a flash memory of the prior art, it is known that the variation of the threshold value is increased due to both a weak program (the disturb) or the like caused by the programming, reading and erasing operations of an adjacent bit and natural leakage (the retention), and consequently, the half-value width (the width of the peak of the bell-shaped variation distribution at the position of a half peak value, as shown in Fig. 3) of the variation distribution of the threshold value corresponding to logic "0" and "1" increases with the lapse of time. The inventors have found that, with the lower level of the power supply voltage of future LSIs, the threshold voltage of the memory cells may exceed the marginal range for the read voltage by the broadening of

the variation distribution with time, thereby to cause a malfunction.

This problem is especially serious in a multi-value memory for storing one memory element with data of a plurality of bits by the difference between the threshold values, because this difference is small for the individual data. In a flash memory, moreover, there is a technical problem for minimizing the processing time and the circuit scale intrinsic to the multi-value memory, because of the erasing and program verifying operations intrinsic to the nonvolatile memory device.

10 SUMMARY OF THE INVENTION

An object of the present invention is to provide a multi-value type nonvolatile memory device which can realize programming, reading and erasing operations of high accuracy performed in a short time while minimizing the increase in the circuit scale.

15 Another object of the present invention is to provide a method of sharpening the shape of the variation distribution of the threshold values, and accordingly, to a nonvolatile memory device capable of stably operating at a low voltage.

Representatives features of the invention to be disclosed herein will be briefly summarized in the following.

(1) At the data programming time, data of a plurality of bits are transformed by a data transforming logic circuit into data (multi-value data) according to the combination of the bits, and the transformed data are sequentially transferred to a latch circuit connected to the bit lines of a memory array. A program pulse is generated according to the data latched in the latch circuit and is applied to

25

a memory element in a selected state, so that it is brought into a state in which it has a threshold value corresponding to the multi-value data. In the data reading operation, the states of the memory elements are read out by changing the read voltage to intermediate values of the individual threshold values and are transferred to and latched in a register for storing the multi-value data, so that the original data may be restored by an inverse data transforming logic circuit on the basis of the multi-value data stored in the register.

(2) After a weak erasing operation of the memory elements in the memory array has been executed, the memory element, which has a threshold value lower than the read level of the word line and higher than the verify level, is detected, and the program is executed such that the threshold value of the memory element may be lower than the verify voltage thereby to narrow the width of the variation distribution shape of the threshold voltage of the memory element which is programmed correspondingly to the individual input data.

According to the aforementioned feature (1), the peripheral circuit scale of the memory array can be suppressed to a relatively small size. In the programming operation, the verify voltage value of the word line is sequentially changed (as will be seen in (1) to (4) of Fig. 3) by a predetermined value in a direction away from the near side of the erasing word line voltage, so that the total number of the program pulses, i.e., the program time, can be made shorter than that of the multi-value flash memory system, in which the verify voltage is set at random, thereby to realize a programming operation performed in a short time.

According to the aforementioned feature (2), on the other hand, the shape of the variation distribution of the threshold voltage of the memory elements, which has been widened due to disturb or retention influences, can be returned to a steep shape substantially identical to that just after the end of the programming operation.

BRIEF DESCRIPTION OF DRAWINGS

Fig. 1 is an explanatory diagram showing one example of the operation of transforming two-bits of data to be programmed in one memory cell and read therefrom into quaternary data of the level to be physically programmed in each memory cell and read therefrom in accordance with the present invention.

Fig. 2 is an explanatory diagram showing one example of the operation of inversely transforming quaternary data transformed by a data transforming logic circuit into the original two-bits of data.

Fig. 3 is an explanatory diagram showing the relations between the quaternary data and the threshold values of the memory cells.

Fig. 4 is a circuit diagram schematically showing one embodiment of a multi-value flash memory according to the present invention.

Fig. 5 is a flow chart showing a programming procedure of the multi-value memory of the embodiment.

Fig. 6 is a timing diagram showing the programming operation waveforms of the multi-value flash memory of the embodiment.

Fig. 7 is an explanatory diagram of the operation waveforms showing the difference between the programming method of the multi-value flash

memory of the embodiment and another programming method.

Fig. 8 is a flow chart showing a reading procedure of the multi-value flash memory of the embodiment.

Fig. 9 is a timing diagram showing the reading operation waveforms of
5 the multi-value flash memory of the embodiment.

Fig. 10 is block diagram showing an example of the entire structure of the multi-value flash memory of the embodiment.

Fig. 11 is a block diagram showing an example of the system construction of an embodiment in which a controller is given a function to
10 transform two-bits data intrinsic to the multi-value memory and quaternary data.

Fig. 12 is a diagram showing the structure of a memory cell used in the flash memory of the embodiment and the voltage state at the programming time.

Fig. 13 is a diagram showing the voltage state at the erasing time of
15 the memory cell used in the flash memory of the embodiment.

Fig. 14 is a diagram showing the voltage state at the reading time of the memory cell used in the flash memory of the embodiment.

Fig. 15 is a block diagram showing an internal power source generator
20 and a switching circuit for selecting and feeding the generated voltage to a word driver or the like.

Fig. 16 is a schematic circuit diagram showing an example of the construction of the word driver.

Fig. 17 is an explanatory diagram showing a method of an embodiment
25 for refreshing the multi-value flash memory.

Fig. 18 is a flow chart showing a refreshing procedure of the multi-value flash memory of the embodiment.

Fig. 19 is a timing diagram showing the operation waveforms at the refreshing time.

5 Fig. 20 is a schematic circuit diagram showing an example of the construction of a sense latch circuit of an embodiment.

Fig. 21 is a schematic circuit diagram showing the state at the data inversion starting time and the operation of the sense latch circuit.

10 Fig. 22 is a schematic circuit diagram showing the state at the data inversion ending time and the operation of the sense latch circuit.

Fig. 23 is a schematic circuit diagram showing the state at the verifying time and the operation of the sense latch circuit.

DESCRIPTION OF THE PREFERRED EMBODIMENT

15 With reference to the accompanying drawings, an embodiment of the invention as applied to a flash memory will be described.

Fig. 1 shows a method of transforming data to be inputted from the outside and stored, and multi-value data to be stored in memory cells, and Fig. 2 shows an inverse transforming method for restoring the original data
20 from the multi-value data.

Fig. 1 shows an example of the transformation method in which any of two bits, i.e., "00", "01", "10" and "11" is to be stored in one memory cell, although the invention is not especially limited thereto. There are four kinds of combinations of the first binary data "a" and the second binary data "b" in Fig.
25 1(1), and these individual combinations are transformed into four kinds of

quaternary data having 0, 1, 2 and three "1" in the four bits, by executing three kinds of logic operations ($a \text{ NAND } b$), ($\text{NOT } b$) and ($a \text{ NOR } b$), as shown in Fig. 1(2).

Here, if the memory elements whose number is equal to that of the number of "1s" as a result of the aforementioned operation are subjected to a programming operation, i.e., the application of program pulses, they will have four types of threshold values, as shown in Fig. 1(3), according to the number of programming operations, so that two-bits of data can be programmed in one memory cell. The states of the changes in the threshold value distributions of the individual memory elements are shown in Fig. 3 when the data "00", "01", "10" and "11" are to be stored in the same number in a plurality of memory elements in the memory array.

Fig. 2 shows the data reading principle. By changing the read voltages of word lines at three stages (intermediate values of the individual threshold value distributions of Fig. 3), three kinds of data "c", "d" and "f" can be sequentially read out from one memory cell. Then, one (a) of the programmed two-bits data can be restored by executing a logic operation $(d^* \text{ NAND } f) \text{ NAND } c^*$ on the read-out data. Of the read-out data, moreover, the data d are identical, as they are, to the program data b. Incidentally, symbols d^* and c^* designate the inverted signals of the data d and c.

Fig. 4 shows one specific example of the circuit construction for the transformation of multi-value data and the inverse transformation, as shown in Figs. 1 and 2.

At the data programming time, the data of $2n$ -bits bit length, fed from the outside to the multi-value flash memory, are serially stored through a

switch SW1 in two binary data registers REG1 and REG2 having a data width of n-bits. At this time, the switch SW1 is changed by the output of a flip-flop FF1 driven by a clock signal CLK1 fed from the outside, although the invention is not especially limited thereto, and a clock signal CLK1', produced in a frequency divider DVD by dividing the clock signal CLK1 and having a frequency twice as large as the clock signal CLK1, is fed through a change-over circuit CHG. In synchronism with this clock signal CLK1', the binary registers REG1 and REG2 are shifted, so that the input data are alternately latched bit by bit in the binary data registers REG1 and REG2.

The data "a", latched in the first binary register REG1, and the data "b", latched in the second binary register REG2, are shifted in synchronism with the clock signal CLK2 which is fed from an internal clock generator 30 through the change-over circuit CHG, fed bit by bit to a data transforming logic circuit 11 for the operations of Fig. 1(2), and then sequentially transferred after a predetermined logic operation through a switch SW2 to a sense latch circuit 13 having an n-bits length and disposed on one side of a memory array 12, so that they are programmed in the memory cells of the memory array 12. These programming operations will be described later in more detail.

The aforementioned change-over circuit CHG is switched, by a control signal coming from a sequencer 18 for controlling the memory inside, to feed the clock signal CLK1' to the binary registers REG1 and REG2 at the data input time and the clock signal CLK2 from the clock generator 30 to the binary registers REG1 and REG2 at the time of data transfer with the sense latch 13.

The aforementioned data transforming logic circuit (the data programming operation circuit) 11 is constructed to include: a NAND gate G1

which is allowed to receive at the individual input terminal the data a and b in the binary data registers REG1 and REG2 and to perform the operation (aNANDb) and a NOR gate which is also allowed to receive at the input terminal the data a, b and to perform the operation (aNORb); and an inverter G3 which is allowed to receive at its input terminal the data b of the binary data register REG2 and to perform the operation (NOTb). The switch SW2 selects and feeds any of the output signals of those logic gates G1, G2 and G3 to the sense latch circuit 13.

At the data reading time, on the other hand, the read data "c", having appeared on a bit line in response to the setting of one word line in the memory array 12 to the read voltage level, are amplified and latched by the sense latch circuit 13 and are serially transferred through a switch SW3 to the binary data register REG1 in synchronism with the internal clock signal CLK2.

Next, the data "d", read out to the sense latch circuit 13 by changing the read voltage level, are serially transferred to the binary data register REG2 through the switch SW3. Moreover, the data "f", read out to the sense latch circuit 13 by changing the read voltage level, are serially transferred to an inverse transforming logic circuit 14 through the switch SW3. At this time, the binary registers REG1 and REG2 are shifted in synchronism with the clock signal CLK2.

Here, the period of the clock signal CLK2 at the data reading time may be shorter than that of the clock signal CLK2 at the data programming time. The clock signal CLK2 can be generated which has a period determined by the clock generator 30 in accordance with the control signal from the sequencer 18. The change in the word line reading level is also changed

according to the control signal from the sequencer 18.

The inverse transforming logic circuit (the data reading operation circuit) 14 is constructed to include: an inverter G11 for receiving the data outputted from the binary data register REG2; a NAND gate G12 for directly receiving at its input terminals both the output of the inverter G11 and the data transferred from the sense latch circuit 13; a delay circuit DLY for delaying the data outputted from the binary data register REG1 and transmitting the delayed data at a predetermined timing; an inverter G13 for inverting the signal coming from the delay circuit DLY; and a NAND gate G14 for receiving the output of the inverter G13 and the output of the NAND gate G12. The logic operations $(d \cdot \text{NAND } f) \cdot \text{NAND } c^*$, shown in Fig. 2, are executed for the read-out data c and d, latched in the binary data registers REG1 and REG2, and for the read-out data f, transferred directly from the sense latch circuit 13. These operation results are outputted through the switch SW1 to a data input/output terminal I/O.

At the same time that the one-bit data are thus outputted, the binary data register REG2 is shifted so that one bit of the data "d" (=b) latched are outputted. At this time, the shift operation of the binary registers REG1 and REG2 are synchronized with the clock signal CLK2. Next, the next bits of the data "c" and "d" are read out again from the binary data registers REG1 and REG2, and the logic operations $(d \cdot \text{NAND } f) \cdot \text{NAND } c^*$ are executed for the next one bit of the read-out data "f" which are directly transferred from the sense latch circuit 13. By repeating operations similar to the aforementioned ones, the data "a" and "b", having been backward transformed and restored to the original two bits, are outputted to the outside from the data input/output

terminal I/O.

Incidentally, the data "a", inversely transformed by the inverse transforming logic circuit 14, do not necessarily need to be instantly outputted to the input/output terminal I/O, as described above. Alternatively, the data "a" inversely transformed may be first latched in the binary data register REG1 and then outputted to the input/output terminal I/O alternately with the data in the binary data register REG2 after all bits have been inversely transformed. In this alternative case, a one-bit latch circuit may desirably be provided in place of the aforementioned delay circuit DLY.

As a result, the data "c" in the binary register REG1 can be read out bit by bit and logically operated with the data "d" and "f", and the results can be simply programmed in the original bit positions in the binary data register REG1. The shift operations of the binary registers REG1 and REG2, when the inversely transformed data are outputted to the outside after they are once latched in the binary registers REG1 and REG2 can be synchronized with the clock signal CLK1 coming from the outside.

The flash memory of this embodiment is equipped, although the invention is not especially so limited, with: a command register 16 for latching the command fed by an external CPU or the like; a command decoder 17 for decoding the command latched in the command register 16; and a sequencer 18 for sequentially producing and outputting the control signals for the individual circuits, such as the aforementioned switches SW2 and SW3, to execute the processings corresponding to those commands on the basis of the decoded results of the command decoder 17. The flash memory thus constructed decodes the commands, when fed, and executes the

corresponding processing automatically. The aforementioned sequencer 18 is constructed to include a ROM (Read Only Memory) latching a series of micro instruction groups necessary for executing the commands (or instructions), like the control unit of a CPU of the micro program type, so that the micro programs are started when the command decoder 17 generates the leading address of the micro instruction groups, which correspond to the commands, and feeds the address to the sequencer 18.

The detailed programming procedure will be described in the following in accordance with the programming flow of Fig. 5.

First of all, prior to the programming, all the memory cells are batch-erased. As a result, all the memory cells are caused to have the highest threshold value (of about 5 V) and to be brought into the state (as shown in Fig. 3(1)) such that they store "11" as the program data. The batch erase operation is carried out, as shown in Fig. 13, by raising the voltage of the word line to apply a voltage of 10 V to the control gate CG of the memory cell, a voltage of 0 V to the drain through the bit line and a voltage of -5 V to the substrate (the semiconductor region p-well) thereby to inject electrons into a floating gate FG. The batch erase process is executed by programming in the command register 16 the erase command coming from the external CPU to instruct the erase operation.

Incidentally, in Fig. 13 (Fig. 12 and Fig. 14): reference symbol psub designates a p-type semiconductor substrate; pwell designates a p-type semiconductor well region for the base of the memory cell; niso designates an n-type semiconductor isolation region for effecting the isolation from the substrate pseb at the data erasing time (at the negative voltage applying

time); n+ in the surface of the p-type well region pwell designates the source and drain regions of the memory cell; and p+ in the surface of the p-type well region pwell, n+ in the surface of the isolation region niso, and p+ in the surface of the substrate psub designate the contact regions for reducing the resistances of the contacts with the electrodes for applying the potentials to the individual semiconductor regions. In one p-type well region there is formed memory cells which are connected to word lines, say, one hundred and twenty eight word lines, although the invention is not especially limited thereto, so that all of the memory cells formed over one well can be batch-erased. Moreover, the memory cells can be erased in a unit of a word line by rendering the word line potential selected (10 V)/unselected (0 V) for the memory cells over one p-type well region.

After the end of the batch erase operation, the flash memory is brought into the program mode by programming in the command register 16 of Fig. 4 the program command coming from the external CPU. In this program mode, the program data are inputted at a predetermined timing. Then, the flash memory sends the program data to the binary data registers REG1 and REG2, so that the program data are transferred in units of two bits to the transforming logic circuit 11 and are transformed into quaternary data (at Step S1). The transformations are carried out in the order of aNANDb, NOTb (inversion of b) and aNORb. The transformed data (the first transformation is aNANDb) are transferred to the sense latch circuit 13 (Step S2).

At the next Step S3, it is judged whether or not all the data in the binary data registers REG1 and REG2 have been transferred. If this judgment is YES, a program pulse having a predetermined pulse width is applied to the

memory cell of the bit which corresponds to the value "1" of the X (row) address fed from the external CPU and the Y (column) address outputted from a built-in Y-address counter 33, as shown in Fig. 10, (at Step S4), so that the programming is executed. The programming is carried out, as shown in

5 Fig. 12, by applying a voltage of -10 V to the control gate CG through the word line, a voltage of 5 V from the sense circuit to the drain through the bit line, and a voltage of 0 V to the substrate. Incidentally, at this time, a voltage Vcc (e.g., 3.3 V) is applied to the unselected word line. As a result, fluctuation of the threshold value due to the disturb influence is suppressed.

10 Next, the verify voltage (about 3.5 V for the first time) corresponding to the program level is fed to the word line, which is left in the selected state at the programming time, to read out the data in the memory cell to which the program pulse has been applied. The data "0" is read out as read-out data from the memory cell which has been sufficiently programmed, whereas the

15 data "1" is read out from the memory cell which has been insufficiently programmed. It is, therefore, judged according to the programmed data whether the program has been ended or insufficient. Here, the data of the sense latch circuit 13, which have been programmed, are inverted to "0" (Step S6). Moreover, it is judged whether or not all the latch data of the sense latch

20 circuit 13 take the value "0". If all take "0", the programming in this procedure is ended. If there is any insufficiently programmed memory cell having the latch data "1", the routine is returned from Step S7 to Step S4, so that the program pulse is applied again to the memory cell which is insufficiently programmed to have the value "1". By repeating Steps S4 to S7, the program

25 pulse is repeatedly applied so that the threshold values of all the memory cells

may become lower than the program verify voltage. As a result, the programmed memory cells have a threshold value of about 3.2 V on an average.

When the programming of the desired data in all the memory cells is ended by the aforementioned program verifying operation, all the data of the sense latch circuit 13 will take the value "0", so that the routine advances to Step S8, at which it is judged whether or not the programming operations for all the program levels have ended, that is, whether or not the data "10", "01" and "00" have been programmed. If the judgment is NO, the routine is returned to Step S1, at which quaternary data based on the next operation result (NOTb) are programmed in the memory cells to change the verify voltage of the word line (2.5 V for the second time). As a result of this verification, the programmed memory cells have a threshold value of about 2.2 V on an average. After this, the programming and verification (at a verify voltage of 1.5 V) of the third operation result are executed, so that the programmed memory cells have a threshold value of about 1.2 V on an average, thus ending the programming.

Fig. 6 shows the waveforms of the control clock signal CLK2, the data to be programmed in the sense latch circuit 13, and the potential of the selected word line in the aforementioned programming and program verifying operations. In the first programming, the first operation result (aNANDb) is transferred to the sense latch circuit 13, and the selected memory cell having the latch value "1" is then programmed by the program pulse. Next, a voltage of about 3.5 V, for example, is fed as the program verify voltage to the word line, and it is judged whether or not the programmed data have the value "0".

When the threshold value is higher than 3.5 V, the read-cut data have the value "1" and are found to have been insufficiently programmed, so that the programming operations are repeated till the read-out data have the value "0". Next, the second operation result (NOTb) is transferred to the sense latch circuit 13, so that the programming operation of the desired memory cell is started by the program pulse. The program verify voltage is set to about 2.5 V, and it is judged whether or not the programming is insufficient. If the judgment is YES, the programming is executed again. Finally, the third program result (aNORb) is transferred to the sense latch circuit 13, and a procedure like the aforementioned one is executed. The program verify voltage in this case is about 1.5 V.

In the foregoing embodiment, as described above, the setting of the word line voltage at the three program verify stages is so controlled that the voltage value is sequentially changed (3.5 V ? 2.5 V ? 1.5 V) away from the erase level from the starting point of the level (3.5 V) which has been set at the closest value to the erase level (about 5 V). In the foregoing embodiment, moreover, even the memory cell of which the target threshold value is an intermediate or lowest value (2.2 V, 1.2 V) is programmed simultaneously with the programming of the memory cell whose target threshold value is the highest value (3.2 V), as shown in Fig. 7(3). This is one of the features of the present invention. As a result, the increase in the programming time of the multi-value data can be minimized.

Specifically, in addition to the aforementioned method, a conceivable method for setting the programming and program verifying word line voltage is one in which the setting is changed so as to execute the first programming of

memory cells, as the programming object, having an intermediate threshold voltage (2.2 V) out of the three kinds of threshold voltage, and the second programming of memory cells, as the programming object, having a voltage (3.2 V) higher than the voltage of the first programming or a voltage (1.2 V) lower than that. As shown in Fig. 7(A), alternatively, there can be conceived a method for batch-programming the memory cells having an identical target threshold value. According to these methods, however, it takes a long time for programming, and the time for the charge/discharge to change the word line voltage is increased, so that the time for the program/verify becomes longer than that of the present embodiment.

Next, the reading operation of the memory cells will be described with reference to Figs. 8 and 9. The data reading operation is performed, as shown in Fig. 14, by raising the voltage of the word line to apply a voltage of the selected level, such as 3.7 V, 2.7 V or 1.7 V, to the control gate CG of the memory cell or a voltage of 1.5 V to the drain through the bit line. The reading operation is executed by programming the command for ordering a reading operation in the command register 16.

When the reading operation is started, the read level is set at first to the highest level of 3.7 V to energize the word line (at Step S11). Then, in the selected memory cell, data will appear on the bit line in accordance with the word line reading voltage level, so that the data are read out by amplifying the bit line level by the sense latch circuit 13 (Step S12). Next, the subsequent steps are different depending upon whether the reading is the first, second or third reading (Step S13). Specifically, when the reading is the first reading, the read data in the sense latch circuit 13 are transferred to the binary data

register REG1 (Step S14).

When the transfer of all the read data in the sense latch circuit 13 has ended, the routine returns from Step S15 to Step 511, at which the second data reading operation is executed by setting the read level to 2.7 V to

- 5 transfer the read data to the binary data register REG2. At the end of the second data read and transfer, the third data reading operation is performed by setting the read level to 1.7 V, and the routine moves from Step S13 to Step S16, at which the read data are transferred directly to the inverse transforming logic circuit 14. Moreover, the data, latched in the binary data
- 10 registers REG1 and REG2, are individually transferred bit by bit to the inverse transforming logic circuit 14, in which there is executed a logic operation for transforming the quaternary data into two bit data (Step 517). Moreover, the foregoing procedure (Steps 16 to 18) is repeated to end the reading operations till the transfer and transformation of all the data in the sense latch
- 15 circuit 13 are ended. The data transformation is effected by executing the operation of Fig. 2.

Fig. 9 shows the timings of the control clock CLK2 in the reading operation according to the aforementioned procedure, the data to be transferred from the sense latch circuit 13, and the read level of the word line.

- 20 When the read command and the address are fed from the outside, the reading operation is started to set the first read level (3.7 V) at first thereby to activate the word line, so that the data will appear on the bit line. The data "c", having appeared in response to the first word line level 3.7 V, are read out by the sense latch circuit 13 and are transferred to the first binary data register
- 25 REG1 having a data width equal to n bits, which represents the data length of

the sense latch.

Next, the data "d", produced by lowering the word line level by a predetermined value to the second read level 2.7 V, are transferred to the second binary data register REG2. The data "f", produced by lowering the word line to the third read level 1.7 V, are transferred to the inverse transforming logic circuit 14 so that the aforementioned quaternary data "c", "d" and "f" are changed again to two-bit data and outputted to the outside, such as to the CPU.

Fig. 10 shows the relation between an example of the entire construction of the multi-value flash memory MDFM having on the common semiconductor chip the aforementioned data transforming/inverse-transforming circuit, and a controller CONT connected with the flash memory MDFM. This controller CONT may have only an address generating function and a command generating function for the multi-value flash memory of the present embodiment, so that a general purpose microcomputer can be used.

In Fig. 10, the circuit components designated by the same reference symbols of Fig. 4, have identical functions. Specifically, the symbols REG1 and REG2 designate binary data registers for holding the program data of two bits; the numeral 11 designates a data transforming logic circuit for transforming the held two-bits data into quaternary data; the numeral 12 designates a memory array provided with nonvolatile memory elements having a floating gate, such as a FAMOS in a matrix form; the numeral 13 designates a sense latch circuit for latching the read data and the program data; the numeral 14 designates an inverse transforming logic circuit for transforming the quaternary data read out from the memory array into two-bits

data; the numeral 16 designates a command register for latching the command fed from the controller CONT; the numeral 17 designates a command decoder for decoding the command code held in the command register 16; and the numeral 18 designates a sequencer for sequentially
5 generating and outputting the control signals for the individual circuits in the memories to execute the processings corresponding to the commands.

The multi-value flash memory of this embodiment is equipped with two memory arrays, although the invention is not especially limited thereto, and individual sense latch circuits 13 are provided for the respective memory
10 arrays. These individual sense latch circuits 13 are constructed to simultaneously amplify and latch the data of the memory cells of one line sharing the word line in the memory array, so that the read data latched in two sense latch circuits 13 are selected by a common Y-decoder 15 and transferred bit by bit or in units of a byte to an output register 19. The read
15 data latched in the output register 19 are outputted to the external CPU or the like through a buffer circuit 22. The sense latch circuit 13 of the embodiment of Fig. 4 performs a shift operation during the data transfer and is required to have a function similar to that of the shift register. However, the sense latch circuits 13 can have no shift function by providing a construction, as in Fig. 10,
20 in which the data are selected by the Y-decoder 15 and in which this Y-decoder 15 shifts the selected bit in response to the clock signal.

The multi-value flash memory of this embodiment is constructed to include, in addition to the above specified individual circuits, an all decision circuit 20 for deciding whether or not the data read out from the memory array
25 12 and fed to the sense latch 13 are all "0" or all "1"; a buffer circuit 21 for

fetching external control signals, such as a reset signal RES, a chip select
 signal CE, a program control signal WE, an output control signal CE, a system
 clock SC and a command enable signal CDE indicating whether the input is a
 command input or an address input, all signals being fed from the controller
 5 CONT; a buffer circuit 22 for fetching an address signal and a command
 signal; an internal signal generator 23 for generating a control signal for an
 internal circuit on the basis of the external control signal; an address register
 24 for latching the address which has been held in the buffer circuit 22; a data
 register 25 for latching the input data; X-address decoders 26a and 26b for
 10 decoding the fetched address to generate a signal and for selecting the word
 line in the memory array 12; a word driver 27; an internal power source
 generator 28 for generating voltages required in the chip, such as the
 substrate potential, the program voltage, the read voltage and the verify
 voltage; a switching circuit 29 for selecting a desired voltage from these
 15 voltages in accordance with the operating state of the memory and feeding
 the selected voltage to the main decoder 27 and the like; a clock generator 30
 for generating the internal clock signals (CLK2 and the like); a timer circuit 31
 for counting the clock pulses to give times, such as a program pulse width; a
 status register 32 for indicating the control state of the memory by the
 20 sequencer 16; a Y-address counter 33 for updating the Y-address
 automatically; a false address register 34 for latching the position (address) of
 a false bit; a redundancy comparator 35 for comparing the Y-address and the
 false address; and a relieved address register 36 for storing a relieved
 address to switch the selected memory column when the address coincides.
 25 Moreover, the multi-flash memory of this embodiment is constructed to output

a ready/busy signal R/B* for indicating whether or not the memory can be accessed from the outside.

Moreover, the multi-value flash memory of this embodiment is given a function (hereinafter referred to as the refresh function) to sharpen the bell-shaped variation distributions of the threshold values when the shapes broaden and lower due to the disturb or the retention influences (see Fig. 3). This refresh function is activated when a command is fed from the outside as in the programming or erasing operation. If the refresh command is fetched by the command register 16, the sequencer 18 of the micro program control type is started to effect the refreshing operation. This refreshing operation will be described in detail hereinafter. The signal indicating the decision result of the aforementioned all decision circuit 20 is fed to the sequencer 18. In the refreshing mode, the all decision circuit 20 decides that the read data are all "0". When a signal indicating this decision result is fed to the sequencer 18, the sequencer 18 stops the refreshing operation. At the data erasing time, on the other hand, the sequencer 18 stops the erasing operation if the aforementioned all decision circuit 20 decides that the read data are all "1".

In this embodiment, moreover, there is adopted a predecode system in which the X-address decoder decodes the address signal at the two stages by means of the predecoder 26a and the main decoder 26b. The desired word line is selected, for example, by decoding the more significant three bits of the X-address at first using the predecoder 26a and by controlling the word driver 27 with the predecode signal. By adopting such a predecode system, the unit decoders constituting the main decoder 26b can be arranged in a high integration state according to the word line pitch of the memory array thereby

to reduce the chip size.

Incidentally, the multi-value flash memory of the aforementioned embodiment is equipped on the common silicon substrate, as shown in Fig. 4 and 10, with the function circuits 11 and 14 for transforming two-bits data into quaternary data and vice versa. However, a dedicated controller unit having those functions can be separately provided. In this modification, the flash memory chip need not be provided with the functions intrinsic to the multi-value, so that its chip area does not increase. Another advantage is that a plurality of flash memories MDFM can be connected to a single controller unit CONT and controlled by a bus BUS, as shown in Fig. 11. This controller unit is constructed to have an address generating function and a command generating function in addition to the aforementioned data transforming/inverse-transforming functions.

Fig. 15 shows the internal power source generator 28 for generating the word line voltage and a substrate potential V_{sub} and the switching circuit 29 for selectively feeding them to the word driver 27 and the like, and Fig. 15 shows an example of the construction of the word driver 27. The internal power source generator 28 generates the necessary word line voltages in response to the internal control signals which are generated from the sequencer 18 in correspondence to the various operation modes. The construction of the internal power source generator 28 for generating voltages including the word line voltage and the construction of the switching circuit (the word line voltage switching circuit) 29 for receiving the generated voltages are similar to those of the prior art, except that the kinds of the voltage values of the word line are increased for the multi-value operation.

Specifically, there are four kinds of word line voltages necessary for the binary flash memory of the prior art: the read voltage (2.7 V, 0 V); the program voltage (-10 V, 0 V); the program verify voltage (1.5 V); the erase voltage (+10 V, 0 V) and the erase verify voltage (4.3 V, 0 V). On the contrary, the word
5 line voltages necessary for the multi-value flash memory of the present embodiment are: the read voltage (3.7 V, 2.7 V, 1.7 V, 0 V); the program voltage (-10 V, 0 V); the program verify voltage (3.5 V, 2.5 V, 1.5 V); the erase and erase verify voltages (10 V, 4.3 V, V); and the refresh voltage (-10 V, 10 V, 3.7 V, 3.5 V, 2.7 V, 2.5 V, 1.7 V, 1.5 V, 0 V).

10 The aforementioned switching circuit 29 receives the internal control signals, which are generated by the sequencer 18 and correspond to the various operation modes, and feeds the voltages, generated by the aforementioned internal power source generator 28, to the power terminals P1 and P2 of the word driver 27 which is constructed as shown in Fig. 16.

15 The word driver WDRV of Fig. 16 is a driver used when the word line predecoding method is adopted. Eight voltage selectors VOLS1 to VOLS8 have their inputs connected in common to the output node N1 of a logic selector LOGS1, and eight voltage selectors VOLS9 to VOLS16 have their inputs connected in common to the output node N2 of a logic selector LOGS2,
20 so that the individual voltage selectors may be selected by predecode signals Xp1 and Xp1* to Xp8 and Xp8*. Signals XM and XN together with the predecode signals Xp1 and Xp1* to Xp8 and Xp8* are fed from an address decoder XDCR (26b). At this time, the voltage selectors VOLS1 to VOLS16 have to select and feed the same voltage to the word line as that which is
25 unselected by the other logic selector, unless the operation is selected by the

predecode signal even if either logic selector LOGS1 or LOGS2 corresponding to the voltage selectors outputs the select signal of the select level.

For these operations, separating MOSFETs Q56 and Q57 are switched by the predecode signal. In order that a voltage in the unselected state may be outputted to the word line when the separating MOSFETs Q56 and Q57 are cut off, there are further provided a pull-up MOSFET Q58 and a pull-down MOSFET Q59 which can be switched complementarily with the separating MOSFETs Q56 and Q57 to feed a predetermined voltage to each input of the output circuit INV2.

In Fig. 16, the aforementioned signal XM is deemed to be a three-bit signal for indicating which word lines out of the eight word line groups each including eight word lines is to be selected. The predecode signals Xp1 and Xp1* to Xp8 and Xp8* are deemed to be complementary signals for indicating which word line contained in each word line group is to be selected. According to the present embodiment, the high level of the select signal SEL is the select level, and the high and low levels of each of the predecode signals Xp1 and Xp1* to Xp8 and Xp8* are the select level.

The voltage to be fed to the terminal P1 of the aforementioned word driver WDRV is a voltage Vpp to be used for the erasing, programming, verifying and reading operations, such as 5 V, 4.3 V, 3.7 V, 3.5 V, 2.7 V, 2.5 V, 1.7 V, 1.5 V or 0 V. The voltage to be fed to the terminal P2 is either a voltage Vee to be used for the programming and refreshing operations, such as -10 V or a voltage Vss as the ground potential or the reference potential of the circuit, such as 0 V.

Each of the aforementioned logic selectors LOGS1 and LOGS2 is constructed to include: an inverter INV1 for inverting the signal of the X-decoder-XDCR; a transfer gate TG1 for transmitting or blocking the output of the inverter INV1; and a transfer gate TG2 for transferring or blocking the signal of the X-decoder XDCR.

The aforementioned voltage selectors VOLSI to VOLS16 are made to have identical constructions, each of which is made, as represented by the voltage selector VOLS1, of: an N-channel type pull-up MOSFET Q58 connected between a terminal P3 and the gate of a MOSFET Q52 and switched by the predecode signal Xp1*; and a P-channel type pull-up MOSFET Q59 connected between a terminal P4 and the gate of a MOSFET Q53 and switched by the predecode signal Xp1. The voltage selector VOLSI switches the separating MOSFET Q56 by using the predecode signal Xp1 and the other separating MOSFET Q57 by using the predecode signal Xp1*.

The aforementioned terminals P3 and P4 are fed with the voltage Vcc or Vss.

Next, the operations of the word driver WDRV of Fig. 16 will be described. Table 1 shows the voltages at the terminals and the word line voltages in the individual operation modes. The description of the manner in which to set the program mode, the erase mode and the read mode will be omitted.

[Table 1]

	SELECTED	UNSELECTED	XM	Xp	DE	P4 ●	P1 @	P3 O	P2 O	WORD LINE
ERASE	O		L	H	L	Vcc	Vpp	Vcc	Vss	Vpp
		O	H	H						Vss
		O	L/H	L						Vss
PROGRAM	O		L	H	H	Vss	Vcc	Vss	Vee	Vee
		O	H	H						Vcc
		O	L/H	L						Vcc
READ	O		L	H	L	Vcc	Vcc	Vcc	Vss	Vcc
		O	H	H						Vss
		O	L/H	L						Vss

When the erase mode is specified by the command, the switching circuit 29 feeds the voltage V_{pp} to the terminal P1, the voltage V_{ss} to the terminal P2, and the voltage V_{cc} to the terminals P3 and P4, and the control signal DE is set to the low level.

5 On the other hand, all the bits of the signal XM are set to the low level, so that any of word lines W1 to W8 can be selected. As a result, when the select signal SEL at the select level (the high level) is fed, the node N1 is set to the low level through the inverter INV1 and the transfer gate TG1 so that this low level is fed to the inputs of the individual voltage selectors VOLS1 to
10 VOLS8. When the memory cell to be erased is coupled to the word line W1, only the signals Xp1 and Xp1* of the predecode signals Xp1l and Xp1* to Xp8 and Xp8* are set to the high level and the low level, respectively.

Therefore, the separating MOSFETs Q56 and Q57 of only the voltage selector VOLS1 are turned on, so that the signal at the node N1 is fetched by
15 the voltage selector VOLS1. At this time, both the pull-up MOSFET Q58 and the pull-down MOSFET Q59 of the voltage selector VOLS1 are cut off.

As a result, the signal of the node N1 is fed to the MOSFETs Q52 and Q53 of the voltage selector VOLS1. Then, the MOSFET Q52 of the output circuit INV2 is turned on, and hence the word line W1 begins to be charged by
20 the voltage V_{pp} at the terminal P1. At this time, the low level to be fed to the gate of the other MOSFET Q53 is raised to a low level higher than the initial voltage V_{ss} by the action of the MOSFET Q57, so that the MOSFET Q53 is not completely cut off. However, when the conductance of a feedback MOSFET Q55 is increased with the rise of the level of the word line W1, the
25 voltage of the gate of the MOSFET Q53 is forced to the voltage V_{ss} , and it is

completely cut off.

In the erase mode, therefore, the word line W1, to which is coupled the selected memory cell, is charged to the level V_{pp} .

While the select signal SEL is at the high level, as described above, the
5 predecode signals $Xp1$ and $Xp1^*$ are set to the low level and the high level, respectively, if the memory cell Q1 of the word line W1 is not selected for the erasure. As a result, both separating MOSFETs Q56 and Q57 of the voltage selector VOLS1 are turned off to fetch no signal from the node N1. At this
10 time, both pull-up MOSFET Q58 and pull-down MOSFET Q59 of the voltage selector VOLS1 are turned on.

As a result, the gates of the MOSFETs Q52 and Q53 of the voltage selector VOLS1 are fed with the voltage V_{cc} from the terminals P3 and P4 through the MOSFETs Q58 and Q59. As a result, the MOSFET Q53 of the output circuit INV2 is turned on, so that the word line W1 begins to be
15 discharged to the voltage V_{ss} through the terminal P2. At this time, the high level fed to the gate of the other MOSFET Q52 is lower than the voltage V_{cc} by the threshold voltage of the MOSFET Q58, so that the MOSFET Q52 is not completely cut off. As the level of the word line W1 is lowered by the ON MOSFET Q53, the conductance of the feedback MOSFET Q54 is increased,
20 and the gate of the MOSFET Q52 is forced to the voltage V_{pp} , so that it is completely cut off. In the erase mode, therefore, the unselected word line W1 is discharged to the voltage V_{ss} .

The operation of the word driver WDRV when the program mode or the read mode is specified, will not be described in detail because it is similar to
25 the operation of the aforementioned program mode. However, the word lines

are so driven by the voltages applied to the terminals P1 and P2 from the switching circuit 29 that the voltages, as shown in Figs. 13 and 14, may be applied to the selected memory cells.

Next, the refreshing operation, which is a second feature of the multi-value flash memory of the present invention will be described with reference to Fig. 17. For the multi-value flash memory which is first programmed with the data, the bell-shaped variation distributions of the threshold values are completely separated, as shown in Fig. 17(1). However, the threshold value variations are increased as shown in Fig. 17(2), as the subsequent programming, reading and standby state operations are repeatedly executed.

This is caused by the so-called disturb influence, in which when a memory cell adjacent to a certain memory cell is programmed, this memory cell is also weakly programmed, and by the retention, influence, which is caused by the natural leakage at the standby time. This phenomenon may occur even in an ordinary flash memory for storing only one bit, but may cause a malfunction in the multi-value flash memory in which the intervals between the individual threshold values are narrow, as in the foregoing embodiment.

In the present embodiment, therefore, there is executed a refreshing operation for sharpening the peaks (see Fig. 3) of the variation distributions of the threshold values when the peaks are broadened and lowered.

Next, the procedure of the refreshing operations will be described.

Fig. 18 is a flow chart showing the procedure of the refreshing operation. When the refresh command is inputted from the external CPU or the like, the sequencer 18 is started to begin the refreshing operation

according to the flow chart of Fig. 18. When the refreshing operation is started, a weak erase pulse is applied at first from the word line to all the memory cells which are connected to the selected word line (Step S21). As a result of this application of the weak erase pulse, the threshold values of all the memory cells are slightly shifted to the higher side, as shown in Fig. 17(3). This shift is about 0.2 V, although the invention is not especially limited thereto. Here, the term, weak erase pulse, refers to a sufficiently short pulse that the memory cell threshold value at "10", for example, may not exceed, if added, just the higher read level 3.7 V. The pulse width is experimentally determined according to the amount to be shifted.

At the second stage, the word line voltage is set to the read level (3.7 V) corresponding to the stored data "10" (Step S22) thereby to effect the reading operation. As a result, data are read out according to the threshold values of the individual memory cells (Step S23) and are amplified and latched by the sense latch circuit 13. At this time, the data of the sense latch corresponding to the memory cell having a higher threshold value than the word line voltage is set to "1", and the data of the sense latch corresponding to the memory cell having a lower threshold value than the word line voltage is set to "0". Next, the data of the sense latch are inverted (Step S24). This data inversion can be easily carried out (as will be described hereinafter) by the sense latch circuit having the construction shown in Fig. 20.

Next, the word line is set to a lower verify voltage (3.5 V at first) which is lower than the aforementioned read level (Step S22) so that the judgment of the threshold value is executed (Step S25). As a result, the data of the sense latch corresponding to the memory cell (designated by letter A in Fig.

17(4)) having a lower threshold value than the verify voltage are switched from "C" to "1". On the contrary, the data of the sense latch corresponding to the memory cell (designated by letter B in Fig. 17(4)) having a higher threshold value than the verify voltage are left at "1". These data are judged to be targets to be re-programmed in the present embodiment. This specifies the memory cells which have excessively approached the read level (3.7 V) when the threshold value is shifted to the higher side by the weak erasure at Step S21. Incidentally, the data of the sense latch, which corresponds to the memory cell (designated by letter C in Fig. 17(4)) corresponding to the stored data "11" having the highest threshold value, are left at "0" set by the aforementioned inverting operation. Such an operation can be automatically executed by a sense latch circuit which has the construction shown in Fig. 20 (as will be described hereinafter).

Therefore, the memory cell (designated by letter B in Fig. 17(4)) having the data "1" of the sense latch is re-programmed by setting the program voltage (Step S27). After this, the verification is executed by setting the verify voltage corresponding to the program level (Step S28, S29). When the threshold value becomes lower than the verify voltage, the latch data change from "1" to "0". The programming and verifying operations are repeated to end the refresh procedure of the memory cell having the data "10", till all the latch data change to "0". As a result, the variation distribution (the half-value width) of the threshold value of the memory cell of the data "10" becomes low, as indicated in Fig. 17(5). From that point on, similar refreshing operations are executed, too, for the memory cells for storing the data "01" and "00" (Step S31). In order to make the width of the distribution shape of the threshold

value, the operations of Steps 21 to 31 are repeated to complete the refresh operation (Step S32) when a predetermined number of operations are ended.

Table 2 shows the changes in the latched data of the sense latch circuit which occur sequentially when the memory cells having the threshold values indicated by letters A, B and C of Fig. 17(4), are read out, when the refreshing is executed according to the procedure described above.

[Table 2]

	READ	INVERSION	VERIFY	END
CELL A	0	1	0	0
CELL B	0	1	1	0
CELL C	1	0	0	0

Fig. 19 is a diagram showing the timings at which the refreshing operations are executed. As described above, the causes of enlarging the variations of the threshold values of the memory cells are the disturb influence due to the execution of the weak programming, erasing and reading operations of a memory cell adjacent to a memory cell if this memory cell is programmed or read, and the retention influence due to the natural leakage.

The timings at which the refreshing operations against the fluctuation of the threshold values due to the disturb influence are executed are as follows.

- (1) The refreshing operations are executed when the flash memory is in a standby state (/RES is at the high level) and after a predetermined number of programming/erasing and reading operations are completed.
- (2) The refreshing operations are executed immediately after the reset

signal (/RES) is activated at the resetting time.

(3) The refreshing operations are executed immediately after the reset state is caused by setting the /RES to the low level from the standby state.

(4) The /RES is set to the low level immediately before the power supply is turned 'off, so that the refresh is executed by sensing the off condition.

(5) The refresh is executed after the power supply is turned on to set the /RES to the high level.

As counter-measures for the reduction of the threshold values due to the retention influence, on the other hand, it is conceivable to execute the refresh operation at predetermined periodic intervals in the course of the dummy cycle or in the standby state when the power supply is turned on. All of these refresh timings may be executed, but any one or only some of them also may be executed.

Incidentally, the refreshing operations described above should not be limited to the multi-value flash memory. As it is considered that the power supply voltage of the flash memory will change to a lower voltage, however, an increase in the variation of the threshold value cannot be ignored even in an ordinary flash memory, so that the refreshing operation is an effective function as a counter-measure against the lower power supply voltage of the flash memory.

Fig. 20 shows an example of the construction of the memory array and the sense latch circuit 13. The memory array 12 is an AND type, in which a plurality of memory cells MC (e.g., one hundred and twenty eight for one hundred and twenty eight batch-erasable word lines) are connected in series between a common drain line DL, provided in parallel with the bit lines BL

arranged perpendicularly to the word lines and adapted to output the read signal of the selected memory cell, and a common source line SL. The common drain line DL can be connected to the corresponding bit line BL through the switch MOSFET Q1, and the common source line SL can be connected to the grounding point through the switch MOSFET Q2. The gate control signals for those switch MOSFETs Q1 and Q2 are generated on the basis of the X-address signal and the read/write control signal. The switch MOSFETs Q1 and Q2 are turned on to discharge the bit lines through the ON memory cells by setting the gate control signals to a potential such as V_{cc} (3.3 V) at the data reading time (including the verifying time). At the data programming time, on the other hand, the gate control signal for the switch MOSFET Q1 is set to a potential such as 7 V and turned on because the program voltage (5 V) of the bit lines is transmitted to the drains of the memory cells. At this time, the switch MOSFET Q2 on the common source line SL side is turned off.

The sense latch circuit 13 is constructed of a CMOS differential type sense amplifier SA disposed for each memory column for amplifying the potential difference between the bit lines of the right and left memory arrays. Prior to the reading operation, the bit line of the selected side (on the lefthand side) memory array is precharged to a potential such as 1 V by a precharge MOS (SW21), and the bit line on the opposite side memory array is precharged to a potential such as 0.5 V by a precharge MOS (SW22).

When the word line WL is set to the read level in this precharge state, the bit line retains 1.0 V if the selected memory cell has a high threshold value. However, if the selected memory cell has a low threshold value, an

electric current flows to draw the charge on the bit line, so that the bit line takes a potential of 0.2 V. The potential difference between this potential of 0.2 V or 0.2 V and the potential of 0.5 V of the bit line on the opposite side is detected and amplified by the sense amplifier SA, so that the read data are
5 latched in the sense amplifier SA.

In the foregoing embodiment, as described before, the sense latch (the sense amplifier) corresponding to the bit line connected to the memory cell to be programmed is set to "1", the program pulse (-10 V) is applied to the word line, and then the word line is set to the verify voltage (about 3.5 V for the first
10 time) corresponding to the program level thereby reading the memory cell to which the program pulse is applied. Moreover, the read data "1" are read out from the insufficiently programmed memory cell to the bit line, and a program end or weak program is judged from the data read out, so that the data of the sense latch (the sense amplifier) whose bits are programmed are inverted to
15 "0". In other words, the data "1" are left in the sense latch (the sense amplifier) corresponding to the insufficiently programmed memory cell, so that the program pulse may be applied again to the insufficiently programmed memory cell corresponding to the bit of "1".

In the refreshing operations, too, the data read out to the sense latch
20 are inverted, and the verification is executed to apply the program pulse to the memory cell corresponding to the bit of "1".

The sense latch circuit of Fig. 20 is devised to have an inversion control circuit 30 which is interposed between the sense amplifier and the memory array and composed of four switches SW11, SW12, SW13 and SW14, so as
25 to easily facilitate the inversion of the latch data of the sense amplifier

corresponding to the memory cell, which has been programmed at the
aforementioned programming time, and the narrowing of the memory cell to
which the program pulse is to be applied.

Here will be described the operation of this sense latch circuit.

- 5 Incidentally, the switches SW21 and SW22 disposed on the individual bit lines
BL are switches for precharging the bit lines and are constructed of
MOSFETs, similar to the aforementioned switches SW11 to SW14.

- At the data reading time, the switch SW13 is turned off at first. With the
bit line BL and the sense amplifier SA disconnected from each other, as
10 shown in Fig. 20, the switches SW21 and SW22 are then turned on to charge
the bit line BL on the selected side to a precharge level of 1.0 V.

- At this time, the bit line on the unselected side is charged to the level of
0.5 V. Moreover, the sense amplifier SA turns on the switch SW14 to reset it
and feed it a potential of 0.5 V. At this time, moreover, the switch MOSFETs
15 Q1 and Q2 are turned on by impressing the voltage V_{cc} to their gates.

- Then, any word line WL in the memory array is set to the select level of
3.7 V. Then, the memory cells (e.g., the cell A and B of Fig. 17) having a
lower threshold value than the word line select level are turned on, so that the
bit line BL connected to them is discharged to the level of 0.2 V by the electric
20 current flowing to the common source line SL through the ON memory cells.
On the other hand, the memory cell (e.g., the cell C of Fig. 17) having a higher
threshold value than the word line select level is turned off, so that the bit line
BL connected to it is held at the precharge level of 1.0 V.

- Next, the switch SW14 is turned off to release the sense amplifier SA
25 from the reset state and to activate it, and the switch SW13 on the bit line BL

is turned on to connect the bit line BL to the sense amplifier SA. The power source voltage V_{cc} is fed to the P-MOS side of the sense amplifier SA, and the ground potential (0 V) is fed to the N-MOS side. Then, the sense amplifier SA amplifies the potential difference between the bit lines BL and BL*

5 sufficiently, and the switch SW13 on the bit line BL is turned off. As a result, the sense amplifier SA comes into the state that it amplifies the, level difference between the bit lines on the select and unselect sides and holds the data.

When the latch data of the sense amplifier SA are to be inverted, the
10 switch SW13 is turned off. With the bit line BL and the sense amplifier SA disconnected from each other, as shown in Fig. 21, the switches SW21 and SW22 are turned onto precharge the bit lines BL on the select and unselect sides to the level of $V_{cc}-V_{tn}$ (e.g., $3.3\text{ V} - 0.6\text{ V} = 2.7\text{ V}$). After this, the switches SW21 and SW22 are turned off, and the switch SW11 is turned on.
15 In accordance with the data latched in the sense amplifier SA, the switch SW12 is then turned on if the data are “is”, so that the bit line BL is discharged to the bit line inverting level (0 V). If the data latched in the sense amplifier SA are “Os”, on the other hand, the switch SW12 is turned off, so that the bit line BL retains the level V_{cc} . In short, the inverse level of the latched data of the
20 sense amplifier SA appears in the corresponding bit line BL.

Here, the switch SW14 is first turned on to reset the sense amplifier SA. After this, the switch SW14 is turned off, and the switch SW13 to the bit line BL is turned on to connect the bit line BL to the sense amplifier SA. In the meantime, the supply voltages on the P-MOS side and the N-MOS side of the
25 sense amplifier SA are set to 0.5 V. Then, the supply voltage V_{cc} is fed to the

P-MOS side of the sense amplifier SA whereas the ground potential (0 V) is fed to the N-MOS side, and the switch SW13 on the bit line BL is turned off. As a result, the sense amplifier SA takes the state that it latches the data corresponding to the level of the bit line in the aforementioned data latching state, as shown in Fig. 22.

In other words, the sense amplifiers corresponding to the cells A and B of Fig. 17 latch the high level "1", and the sense amplifier corresponding to the cell C latches the low level "0". These operations are similar to the so-called "program verifying" operation. Hence, the bit line precharge has to be executed only for the portion in which the sense latch is "H". By turning on the switch SW11 to set the bit line precharge voltage (1) to 1 V, therefore, only bit lines BLO and BL1 take the value of 1 V (the bit line BL2 is reset in advance to 0 V).

Next, the switches SW21 and SW22 are turned on while the switch SW13 on the bit line BL is left off, to charge the select side bit line BL to the precharge level of 1.0 V and the unselected-side bit line to the level of 0.5 V. After this, a verify voltage such as 3.5 V slightly lower than the preceding read level (3.7 V) is applied to the selected word line. Then, the memory cell (e.g., the cell A of Fig. 17) having a lower threshold value than the word line selection level is turned on, so that the bit line BL connected thereto is discharged to the level such as 0.2 V.

On the other hand, the memory cell (e.g., the cell B of Fig. 17) having a higher threshold value than the word line selection level is turned off, so that the bit line BL connected thereto retains the precharge level of 1 V. At this time, moreover, since the bit line, which is connected to the memory cell

(e.g., the cell C of Fig. 17) corresponding the data "11" having the highest threshold value, intrinsically retains the low level, i.e., "0", it takes the low level even if it is off when the word line is set to the select level (Fig. 23).

As a result, after the sense latch is reset in this state, the switch SW13
5 on the bit line BL is turned on. Then, the sense amplifier corresponding to the bit line, which is connected to the memory cell (e.g., the cell C of Fig. 17) corresponding to the data "11", and the sense amplifier corresponding to the bit line, which is connected to the memory cell (e.g., the cell A of Fig. 17) having a lower threshold value than the word line select level, retain the low
10 level "0", whereas the sense amplifier corresponding to the bit line, which is connected to the memory cell (e.g., the cell B of Fig. 17) having a higher threshold value than the word line select level, retains the high level "1". In the present embodiment, this data retained by the sense amplifier are used to make a shift to the programming operation to apply the program pulse (-10 V)
15 to the selected word line, thereby lowering the threshold value of the memory cell corresponding to the retained data "1" of the sense amplifier.

After the application of the program pulse, the reading operation is executed by setting the word line again to the selection level. Then, the bit line of the memory cell having a lower threshold value than the word line verify
20 level is changed to the low level, i.e., "0", and the bit line connected to the insufficiently programmed memory cell retains the high level "1". By latching this state by the sense amplifier to execute the programming operation again, only the threshold value of the memory cell, in which the latched data of the sense latch corresponds to "1", is lowered to sharpen the threshold value
25 distribution shape. The data latched by the sense amplifier A are fed to the

aforementioned all decision circuit 20 through both so-called column switch turned on/off by the output signal of the Y-decoder 15 and the common I/O line, and it is judged whether or not they are all "0". If they are all "0", the refresh for the memory cells of the data "10" are ended, and the refresh for the memory cells of the data "01" and "00" is executed.

Incidentally, the re-programming operation of the insufficiently programmed memory cell in the aforementioned program mode is identical to the aforementioned one effected by the sense latch circuit 13 at the refreshing time.

In the foregoing embodiment, as has been described hereinbefore, at the data programming time, data of a plurality of bits are transformed by a data transforming logic circuit into data (multi-value data) according to the combination of the bits, and the transformed data are sequentially transferred to a latch circuit connected to the bit lines of a memory array. A program pulse is generated according to the data latched in the latch circuit and is applied to a memory element in a selected state, so that a threshold value is made to correspond to the multi-value data. In the data reading operation, the states of the memory elements are read out by changing the read voltage to intermediate values of the individual threshold values and are transferred to and latched in a register for storing the multi-value data, so that the original data may be restored by a data inverse transforming logic circuit on the basis of the multi-value data stored in the register. As a result, the following effects can be achieved. The peripheral circuit scale of the memory array can be suppressed to a relatively small size. In the programming operation, the verify voltage value of the word line is sequentially changed by a predetermined

value in a direction away from the near side of the erasing word line voltage so that the total number of the program pulses, i.e., the program time period can be reduced compared to the multi-value flash memory system, in which the verify voltage is set at random, thereby to realize a programming operation performed in a short time.

Moreover, after a weak erasing operation of the memory elements in the memory array is executed, the memory element, which has a threshold value lower than the read level of the word line and higher than the verify level, is detected, and the program is executed so that the threshold value of the memory element may be lower than the verify voltage, thereby narrowing the width of the variation distribution shape of the threshold voltage of the memory element which is programmed according to the individual input data. As a result, the following effect can be achieved. The variation distribution shape of the threshold voltage of the memory elements, which has been widened due to the disturb or the retention influences, can be returned to the steep shape substantially identical to that just after the end of the programming operation.

Although our invention has been specifically described in connection with its embodiments, it should not be limited to the embodiments specifically described but can naturally be modified in various manners without departing from the gist thereof. In the foregoing embodiments, for example, the quaternary data are stored by setting the threshold value of one memory cell at four stages, but these threshold values can be set to three stages or five or more stages.

In the embodiments, on the other hand, the inversion of the read data

at the refreshing time and the narrowing of the memory cells, in which the read data are to be programmed, can be effected by using only the sense latch circuit. Despite this construction, however, there may be provided a register for latching the read data and a logic circuit for narrowing the memory cell to be programmed, by performing a logic operation, e.g., by inverting the content of the register.

In the embodiments, moreover, the three kinds of operations, as shown in Fig. 1(2), are executed as a transformation of the two-bits data into the quaternary data and vice versa. However, the logic operation should not be limited to those of Fig. 1 but may be any logic operation as long as data having different numbers of bits of "1" can be resultantly obtained. Furthermore, the operation for data inverse transformation should not be limited to those of Fig. 2 but may be any operation as long as the original two-bits data can be restored, and the number of operations should not be limited to one but may be two or more.

The programming method for each memory cell should not be limited to that of the embodiment in which the threshold value is lowered by the program pulse after it has been first raised for the erasure, but may be the one in which the threshold value is raised by the program pulse. In the embodiment, moreover, the threshold value is changed by programming the memory cell corresponding to the sense latch latching the data "1". However, the threshold value may be changed by programming the memory cell which corresponds to the sense latch latching the data "0".

The description thus far made is directed mainly to a batch-erase type flash memory to which our invention is applied and which is the field of

application of its background. However, the present invention should not be limited thereto but can be applied generally to a nonvolatile memory device having FAMOSs as its memory elements and further widely to a semiconductor memory device which is equipped with memory cells having a plurality of threshold values.

According to the present invention, as has been described hereinbefore, it is possible to realize a multi-value type nonvolatile memory device which can carry out programming, reading and erasing operations of high accuracy which are performed in a short time period while minimizing the increase in the circuit scale, and a nonvolatile memory device capable of sharpening the shape of the variation distribution of the threshold values of memory elements while stably operating at a low voltage.